March 2011

WORKING PAPER SERIES

2011-ECO-01

Optimal productive size of hospital's intensive care units

Hervé Leleu
CNRS-LEM (UMR 8179), IESEG School of Management

James Moises
Tulane University, USA

Vivian Valdmanis
University of the Sciences in Philadelphia, USA

IESEG School of Management
Catholic University of Lille
3, rue de la Digue
F-59000 Lille
www.ieseg.fr
Tel: 33(0)3 20 54 58 92
Fax: 33(0)3 20 57 48 55

Optimal productive size of hospital's intensive care units

Hervé Leleu, Ph.D.
CNRS/LEM and IÉSEG School of Management
3, rue de la Digue, FR-59000 Lille
Phone +33/320545892 Fax+33/320574855
h.leleu@ieseg.fr

James Moises, MD
Department of Emergency Medicine, Tulane University
1430 Tulane Avenue, New Orleans, LA USA 77112
Phone +1(504) 988-5623 Fax +1(504) 282-4668
jmoises@tulane.edu

Vivian Valdmanis, Ph.D.
Department of Health Policy and Public Health,
University of the Sciences in Philadelphia
600 S 43rd Street, Philadelphia, PA USA 19104
Phone +1 (215) 596-7613 Fax+(215) 596-7614
v.valdma@usp.edu

Keywords: Hospital, Intensive Care Units, Returns to Scale, Optimal Size
JEL Classification:

Funding Source: N/A
Conflict of Interest: N/A

ABSTRACT

Optimal productive size of hospital's intensive care units

Authors of past studies focusing on returns to scale in hospitals proffered mixed results. These seemingly contradictory findings have probably arisen due to different methodological approaches (parametric or non parametric), different aggregation levels of analysis (hospital/department/units), nature of data (quantity data or economic values) but also technological improvements operating in hospitals and case mix adjustment to account for the severity of patients' conditions.

In this paper, we apply a new approach to determining returns to scale for multi-output homogenous technologies. Our approach is characterized by 1) a non parametric approach based on quantity data allows us to avoid assumptions on cost minimization or profit maximization behavior of hospitals, on relevancy of economic values for hospitals (costs, revenues and prices) and on a priori specification of the health care production function; and 2) an analysis of optimal productivity size at both the disaggregated level of intensive care units and at the aggregated hospital level. The methodological advantage is that we can unambiguously define increasing returns to scale which is lacking in more traditional non-parametric approaches because of the convexity assumption imposed earlier.

We apply the methodology to intensive care units (cardiac care (CICU), medical/surgical care (MSICU), pediatric care (PCIU) and neonatal care (NICU) which are operating in 235 general short term hospitals of Florida state in 2005. We also consider the hospital level by analyzing the general activity of the hospitals in our population.

To summarize our findings, we find that 60% of intensive care units are operating at increasing returns to scale, 10% are operating at optimal productive size and 30% are characterized by decreasing returns to scale. In average intensive care units operate 40% under the optimal size. The policy implication of this result should be an increase of the size of all types of intensive care units to meet productivity gains. The picture is completely reversed at the aggregate hospital level. Here decreasing returns to scale prevail for 65% of hospitals while only one fourth are operating at increasing returns to scale. In average hospitals' number of beds should decrease by 40% to reach the optimal productivity size. One policy solution may include reallocating resources from general beds to the more specialized beds.

Introduction

Whereas there are many units/departments operating in a hospital, one that has not been adequately evaluated are the intensive care units (ICUs). These units have both cost and quality implications. From the cost perspective, ICUs operating in the United States typically absorb 20% of total hospital costs but contain less than 10% of total beds [1]. It has also been argued that larger ICUs report better patient outcomes due to the volume effect [2,3]. Beyond this quality/volume effect, one main reason for interest in the optimal size of ICUs is their potential scale economies. Increasing/decreasing the size of ICUs or merging/disaggregating different types of ICUs could reduce inefficiencies and lower costs. The same arguments also prevail at the hospital level. Derivation of the optimal size of hospitals and the measure of economies of scale was analyzed, partly to justify mergers among hospitals or to show that the most productive scale size was for medium hospitals [4,5,6]. The main point of the literature on this topic is the lack of consensus among previous authors determining economies of scale, especially from a cost perspective.

The research question we address in this paper is: Does the measure of returns to scale vary at the unit level, in this case a variety of intensive care units or at a more aggregated level such as a hospital? We ask this question because what may be confounding previous studies is that increasing returns to scale may prevail at the disaggregated level, while constant or decreasing returns to scale may prevail at a more aggregated level. Therefore, the opportunity to measure specific scale economies at different hospitals' units is crucial to address the question of optimal productive size and to manage a fair allocation of resources among units. Our approach is characterized by 1) a non parametric primal approach based on quantity data which allow to avoid assumptions on controversial economic behavior of hospitals, on relevancy of economic values for hospitals (costs, revenues and prices) and on a priori specification of the health care production function; and 2) an analysis of optimal productivity size at both the disaggregated level of intensive care units and at the aggregated hospital level. The methodological advantage is that we can unambiguously define increasing returns to scale which is lacking in more traditional non-parametric approaches because of the convexity assumption imposed earlier.

In order to address this issue and to solve the methodological problems, we use a new "semi-parametric" approach introduced by Boussemart, Briec, Peypoch and Tavéra [7]. While this approach has been theoretically defined, it has not yet been applied in an empirical work. Following Boussemart et al. [7] we refer to the notion of α-returns to scale to deal with strictly increasing and decreasing returns for multi-output homogenous technologies. This allows testing for returns to scale in the hospital industry in a rigorous way at various level of aggregation: at the services (unit) level, department level or at the aggregated hospital level.

Our application relies on a relatively rich data set: 235 general short term hospitals which are operating in Florida in 2005. By having input and output data by type of ICU including cardiac care units (CICUs), medical/surgical ICUs (MSICUs), neonatal intensive care units (NICUs) and pediatric intensive care units (PICUs) we can directly test whether returns to scale are the same among different services of a hospital. We also have data on the number of beds by service (the typical unit of scale, [8]), labor inputs including full time equivalency (FTE) nursing input, staff and residents. For outputs we have patient days by ICU type. We also have the same data at the hospital level as well as the case mix index of the severity of patients' conditions for 160 of the 235 hospital. A particular benefit of using these data is that we can compare optimal productive size at various ICUs level and at the hospital level. This makes our approach particularly appealing as an addition to the literature.

In the next section of the paper, we present background on hospital and intensive care units economies of scale followed by a more complete description of the methodology used here. In section four the data and results are presented. We close the paper in section five with a discussion of the results and possible policy implications.

Background

Authors of past studies focusing on returns to scale in hospitals proffered mixed results. These seemingly contradictory findings have probably arisen due to different methodological approaches (parametric or non parametric), different aggregation levels of analysis (hospital/department/units), nature of data (quantity data or economic values) but also technological improvements operating in hospitals and case mix adjustment to account for the severity of patients' conditions.

In some studies, researchers have reported that economies of scale are exhibited at hospitals with between 200-300 beds [4,5]or 10,000 discharges [6]. Studies that included hospital costs over time demonstrated changes which included larger hospitals became more scale efficient [9,10]. Preyra and Pink [11] also warn that economies of scale may be sensitive to capital changes leading to different products provided or changes in the case mix of patients treated. Carey [12], Yafchak [8], Lindrooth et al. [13] and Wilson and Carey [10] also pointed out that these changes incur high fixed and/or transaction costs, and if these changes lead to more output being produced then returns to scale are affected. The main policy reason for interest in measuring economies of scale was to justify mergers among hospitals especially if these activities reduce inefficiencies and lower costs. Conversely there could also be some rationale to produce at decreasing returns to scale. There is an agreement among earlier works that larger hospitals, while not producing at economies of scale, may trade off costs with quality. Specifically, the notion of "reservation quality" proposed by Joskow [14] who hypothesized that hospitals will purposely maintain extra beds in case of a surge in admissions, even with additional costs due to decreasing returns to scale. Others have argued that increasing the size of a hospital is a necessary condition for increased volume which in turn results in higher quality of care [15, 12]

Given the issues presented above, there have been suggestions for resolving the estimation issue. Specifically, Wilson and Carey [10] argue that the previous problems with estimating economies of scale in hospitals are because the parametric approach is misspecified. Due to the unique nature of hospitals, i.e., a dominance of non-profit ownership, production of multiple outputs, market failures such as asymmetric and/or incomplete information, and behaviors categorized as other than cost-minimizing or profit-maximizing make determining an a priori functional form impossible. To combat the functional form issue in hospital productivity studies, non-parametric approaches including data envelopment analysis (DEA) have been employed. (See [16] for a review of this literature.) Wilson and Carey [10] also suggest using a non-parametric approach, if the regularity conditions are met, would be better suited for gauging hospital efficiency and productivity. Along this suggestion, Valdmanis [17] applied DEA in order to measure economies of scale by hospitals and hospital markets operating in Florida and found that hospital markets defined as the metropolitan area exhibited constant and decreasing returns to scale a finding suggesting that there are too many beds in a market and decreases in total beds may be warranted. Wilson and Carey [10] used an alternate non-parametric approach by employing kernel analysis and bootstrapping techniques. Even with this more sophisticated approach, these authors report that increasing returns to scale were present for hospital sizes between 72 and 240 beds and decreasing returns to scale for hospitals with bed sizes greater than 360. These results correspond with the other findings of constant returns to scale for hospitals containing between 200-300 beds.

In addition to changing methodology for measuring economies of scale from a parametric to a non-parametric approach, there have been some who have argued that changing the nature of the unit of analysis from the hospital as a whole to individual units [18,11]; as long as the units operating within a hospital can be designated into broad 'iso-resource categories' to meet aggregation conditions [11].

Determining economies of scale has been even more tenuous in the area of ICUs than that of hospitals. Jacobs et al.,[2] found an relationship between lower costs per patient and economies of scale but attributed this finding to the use of fewer inputs rather than lower prices. Gyldmark [19] counters that the costs of ICUs are unknown and if prices are used, then there is a risk of confounding actual resource use with an arbitrary allocated cost based on a hospital's accounting methods. Tsekouras et al., [20] assessed the role of technology in Greek ICUs and efficiency using DEA. These authors found that technical efficiency improved with adoption of new technology combined with medical personnel, scale efficiency remained unchanged, but only analyzed ICUs in general. This may have been an issue since as Edbrooke et al.[1] point out, the diversity of patients in an ICU disallows valid comparisons among ICU units to either determine costs or economies of scale. The problem associated with economies of scale may be amplified if units are typically homogenous in size but not in patient characteristics. The first problem was faced by Jacobs et al. [2] because they did not include large ICUs so that the full range of economies of scale could not be ascertained. The second problem encountered by earlier authors was that patients in ICUs were diverse and the unavailability of precise measurement by a diagnosis did not capture disease severity [1,19,2].

Methods

As suggested by Wilson and Carey [10] and Burgess [18] we use a nonparametric approach which has benefits over typical cost functions parametrically estimated because we do not impose an a priori functional form that may be misspecified but is well behaved in terms of assumptions for determining returns to scale. Before presenting the specific technologies, some introductory remarks regarding returns to scale are appropriate. First, the level of analysis is crucial. Should measuring returns to scale be done at the unit level or at more aggregated levels such as departments or units or the hospital level? We take the approach in this paper to do both – unit specific returns to scale and the hospital level returns to scale which is appropriate for policy and managerial purposes.

Second, it is difficult to test for returns to scale with traditional econometric estimations because of the specificity of the hospital industry. This difficulty may arise since prices for inputs and outputs are often either not available or are not representative of the true costs hospitals face. Aside from the problem of accurate costs and pricing, hospitals are multi-product in nature thereby possibly limiting the ability of stochastic functions to accurately measure hospital productivity. Furthermore, the impossibility of applying duality theory including non-convexity and the too stringent assumption of economic behavior – profit maximization or cost minimization requires using an alternative methodological approach such as DEA. It is because of this contextual issue, that other frontier estimation technique such as the Stochastic Frontier Approach is also not appropriate to use here in the measuring of economies of scale.

Rather than using econometric approaches, the nonparametric approaches such as DEA have been used in the past but they fail to include increasing returns to scale correctly. DEA essentially assumes convexity and if the origin belongs to the production set (zero outputs require zero inputs) only constant and decreasing returns to scale can be modeled. Increasing returns to scale is possible to measure only if a fixed cost is introduced but this alters the homogeneity assumption on production technologies. As demonstrated by Lau [21], one can connect directly returns to scale to a homogenous technology. Therefore, a proper analysis of returns to scale must rely on the homogeneity assumption. A complete characterization of homogenous technologies is given by Färe and Mitchell [22], implying that the estimation of increasing returns to scale may not be properly defined in traditional DEA nonparametric models.

In order to address this issue and solve the methodological problems, we use a new semi-parametric approach introduced by Boussemart et al. [7] and followed up by Boussemart, Briec, and Leleu [23]. Following Boussemart et al. [7] we refer to the notion of α-returns to scale so that we may apply strictly increasing and decreasing returns to scale for homogenous multi-outputs technologies. Homogenous technologies are estimated by varying the α parameters in a grid search approach which than selects the “best” technology for each observation based on a “goodness-of-fit” criterion. This approach allows for the testing for returns to scale for each observation in a rigorous way. We next introduce our technology definitions.

The production technology transforms inputs $x=(x_1,\cdots,x_n)\in R_+^n$ into outputs $y=(y_1,\cdots,y_p)\in R_+^p$ under the technology T:

$$T = \{(x,y)\in R_+^{n+p} : x \text{ can produce } y\} \quad (1)$$

As required by Wilson and Carey [10] that any nonparametric technology described adhere to proper axioms we impose the following:

- T1: $(0,0)\in T$, $(0,y)\in T \Rightarrow y=0$ i.e., no free lunch;
- T2: the set $A(x)=\{(u,y)\in T : u\le x\}$ of dominating observations is bounded $\forall x\in R_+^n$, i.e., infinite outputs cannot be obtained from a finite input vector;
- T3: T is closed;
- T4: For all $(x,y)\in T$, and all $(u,v)\in R_+^{n+p}$, we have $(x,-y)\le(u,-v)\Rightarrow(u,v)\in T$ (free disposability of inputs and outputs).

So far, the production technology and axioms are the same as in the case of the more familiar DEA approach.

We extend the class of technologies to homogenous multi-output technologies. A production technology T is said to be *homogenous of degree* α if for all $\beta>0$:

$$(x,y)\in T \Rightarrow (\beta x, \beta^{\alpha} y)\in T. \quad (2)$$

Lau [20] termed these technologies "almost homogenous technologies of degree 1 and α" for all $\beta>0$. A complete characterization is given by Färe and Mitchell [22]. Obviously, CRS corresponds to $\alpha=1$ while strictly increasing returns corresponds to $\alpha>1$ and strictly decreasing returns corresponds to $\alpha < 1$. Boussemart et. al. [7] termed this property of the technology α *-returns to scale*.

We further propose a nonparametric model of production technologies for which the α -returns to scale technology can be calculated by solving non-parametric DEA models. Let us consider a set of J observations $A = (x_1, y_1), ..., (x_J, y_J) \in R_+^{n+p}$. We denote $J = \{1, \cdots, J\}$. The production technology can be estimated by enveloping observations. Under the usual DEA framework, the production set for constant returns to scale is defined as:

$$T_{CRS} = \left\{ (x, y) \in R_+^{n+p} : x \geq \sum_{j \in J} \lambda_j x_j, y \leq \sum_{j \in J} \lambda_j y_j, \lambda \geq 0 \right\} \tag{3}$$

While its extension to variable returns to scale is given by:

$$T_{VRS} = \left\{ (x, y) \in R_+^{n+p} : x \geq \sum_{j \in J} \lambda_j x_j, y \leq \sum_{j \in J} \lambda_j y_j, \sum_{j \in J} \lambda_j = 1, \lambda \geq 0 \right\} \tag{4}$$

The technology in (3) and (4) corresponds to the traditional CRS case treated respectively in Charnes, Cooper and Rhodes [24] and Banker, Charnes, and Cooper [25]. We note here that the α *-returns to scale* differs from the DEA calculation for returns to scale because there is no convexity constraint, as would be the case when using traditional DEA definitions. We note that DEA is an appropriate technique to apply to determine whether diseconomies of scale exist, but is relatively weak in discerning increasing returns to scale, since increasing returns to scale are limited to the portion of the best practice frontier associated with a fixed cost. While DEA assumes convexity, it must exclude the inactivity strategy (no input/no output) from the set of feasible production plans to model increasing returns. This is a major drawback in the traditional approach. Therefore, by relaxing the convexity assumption and by including the origin in the set of feasible production plans, we can more accurately describe increasing as well as constant and decreasing returns to scale.

As we extend the class of technologies to homogenous multi-ouput technologies, we also use a more general class of technologies introduced by Färe, Grosskopf and Njinkeu [26] and adapted by Boussemart et. al. [7] to α -returns to scale. It is based on a traditional *Constant Elasticity of Substitution* (CES) applied to the input side and on a *Constant Elasticity of Transformation* (CET) formula which characterized the output part. We can now introduce the $CES-CET$ model by considering the following set:

$$T_{\gamma,\delta} = \{(x,y) : x \geq (\sum_{j\in J} \lambda_j x_j^{\gamma})^{1/\gamma}, y \leq (\sum_{j\in J} \lambda_j y_j^{\delta})^{1/\delta}, \lambda \geq 0\} \tag{5}$$

The technology in (5) generalizes the CRS technology case (3) to a more general class of homogenous technologies. It is obvious to see that $T_{CRS} = T_{1,1}$. $T_{\gamma,\delta}$ satisfies T1-T4 and by considering $\alpha = \gamma/\delta$, $T_{\gamma,\delta}$ satisfies α -returns to scale as defined in (2).

From the technology (5), we can derive the following input distance functions:

$$E^I(x_k, y_k) = \min_{\theta^I,\lambda\geq 0}\{\theta^I : \theta^I x_k \geq (\sum_{j\in J} \lambda_j x_j^{\gamma})^{\frac{1}{\gamma}}, y_k \leq (\sum_{j\in J} \lambda_j y_j^{\delta})^{\frac{1}{\delta}}\} \tag{6}$$

In (5) and (6) γ and δ are a priori parameters but finding 'optimal' values for these parameters can be done by applying a goodness-of-fit method. In the application used here, γ and δ vary from 0.05 to 4 at intervals of 0.05 and therefore the linear program (6) is computed 1600 times for each observation. We therefore select the "best" technology as the one associated with the maximal value of $E^I(x_k, y_k)$. From this, we derive the optimal value for γ and δ, and for $\alpha = \gamma/\delta$. We can also describe this relationship using the traditional economic model of a U-shaped average cost curve. As δ – output increases at a greater rather than γ – inputs, this is analogous to the downward slope of the average cost curve, i.e., IRS. When δ equals γ, then we are at the minimum point of the U shaped average cost curve, i.e., CRS. As inputs increase faster than outputs, δ is less than γ, we move to the position on the average cost curve beyond CRS, traditionally referred to as DRS.

Data

Data on hospital discharges and capacity come from the Florida Agency for Health Care Administration. These include staffing, beds, utilization, revenues and expenses by department and in total on 235 general short term hospitals. This is essentially the same information that is reported in the Medicare Hospital Cost Report Minimum Data Set, which is widely-used by researchers. Finally, we get information on case-mix index from Solucient, Inc. which is computed based on DRG pairs derived from Medicare data, and is not strictly a function of patient length of stay. CMI also incorporates a variety of risk factors including severity of illness, complications, co-morbidities and age.

From these data, we conduct our analyses for cardiac intensive care units (CICUs), medical/surgical intensive care units (MSICUs), pediatric intensive care units (PICUs), and neonatal intensive care units (NICUs). An argument for disaggregating ICUs by specific type is in order to account for the diversity of patients within each unit. For each types of intensive care we select data from the 235 hospital population with the two primary criteria: number of days greater than 0 and occupancy rate lesser than or equal to 100%. Secondarily since very few CICUs, MSICUs and NICUs have residents and medical students, we omitted them from our list of inputs. Conversely, half of the PICUs had medical residents providing care and therefore we include them in only in this case. Finally, at the hospital level we select observations for which discharges and capacity information as well as case-mix index are present.

We include four inputs for the evaluation of intensive care units: full time equivalency (FTE) staff, FTE nurses hours; other expenses in dollars, and the number of beds; in the case of PICUs we also include medical residents. Since the staff and personnel as well as beds are specific to each type of care provided, we can assess each unit separately without having to worry that the resources are distributed among other units. For the output we include only the number of days that reliably reflects the level of activity each of the intensive care units under the assumption that patient case-mix is relatively homogenous by the definition of critical care. For the hospital level we also consider the case-mix index to control for the heterogeneity of case mix severity among patients.

We begin by presenting the descriptive statistics of the inputs and outputs (as well as activity indicators) for the different intensive care units and the hospital level (Table 1). We note that there is a wide variation in the size and utilization across the CICUs and that they are typically small (average of 17 beds) with a relatively high occupancy rate (72%). Since typical hospital statistics also present activity indicators, in addition to occupancy rate we note the nurse hours/bed ratio demonstrating that there are, on average, 18.3 FTE nurse hours per bed ranging from 5.8 to 30.7.

Regarding the descriptive statistics for MSICUs, we find similar results as with the descriptive statistics of the CICUs examined. There is variability in terms of both activity indicators ranging from 4% to 100% in occupancy rates and 1.9 to 32.5 FTE nurse hours per bed. However, care is more costly, on average, vis-à-vis the CICUs, but this finding may be due to the fact that

MSICUs care may be less predictable in terms of equipment and personnel required by each patient. For PICUs, on average, these units are smaller than the adult based care units, and the occupancy rates and nurse hours per beds are similarly less than the findings for the adult units. These units are also less costly and are in use fewer days. Neonatal units display the most variability among all the other types of units. The average occupancy rate while at 74% ranges from 10% to 117%. Costs are higher than in the case of PICUs and the size greatly varies from between 4 and 78 beds. Finally, for hospitals as a whole, the mean occupancy rate is much lower (56%) than the averages for any of the intensive care units. Nurse hours per bed are also lower indicating more diversity among all patients requiring intensive nursing attention.

Table 1. Descriptive statistics for intensive care units and hospital level

	Min	Mean	Std	Max
Cardiac ICU (n=48)				
Staff (FTE)	1	25	16	74
Nurses (FTE hours)	35	297	164	798
Other Expenses (1000$)	8.4	772.4	555.6	2862.1
Beds	6	17	10	62
Days	336	4223	2435	13180
Occupancy rate	16%	72%	20%	97%
Nurses hours /beds	5,8	18,5	5,8	30,7
Medical/surgical ICU (n=141)				
Staff (FTE)	1	19	20	180
Nurses (FTE hours)	30	344	294	1684
Other Expenses (1000$)	26.2	1008.0	1024.0	6295.3
Beds	3	20	15	86
Days	244	5160	4237	22384
Occupancy rate	4%	70%	19%	100%
Nurses hours /beds	1.9	17.3	5.6	32.5
Pediatric ICU (n=25)				
Residents (FTE)	0	129	237	847
Staff (FTE)	11	71	52	215
Nurses (FTE hours)	33	212	164	698
Other Expenses (1000$)	101.9	659.4	518.9	2000.6
Beds	3	12	6	23
Days	272	2515	1696	5864
Occupancy rate	19%	56%	17%	83%
Nurses hours /beds	8.4	17.2	6.4	38.8
Neonatal ICU (n=51)				
Staff (FTE)	0	7	6	24
Nurses (FTE hours)	30	353	375	1487
Other Expenses (1000$)	12.2	828.0	789.6	3162.9
Beds	4	22	18	78
Days	250	6551	6698	32851
Occupancy rate	10%	74%	22%	117%
Nurses hours /beds	3.6	14.3	5.2	33.4
Hospital (n=160)				
Staff (FTE)	0	276	298	2412
Nurses (FTE hours)	140	3928	5486	46860
Other Expenses (10000$)	97.6	4762.5	5825.2	41400
Beds	15	324	301	1785
Days	138	69707	74894	497056
Case-mix index	0.869	1.376	0.242	2.083
Occupancy rate	2%	56%	17%	86%
Nurses hours /beds	2.4	11.6	5.3	34.3

Results

Results for intensive care units

We first turn to the main results on returns to scale estimations to all four types of ICUs applying our methodology described below in Table 2.

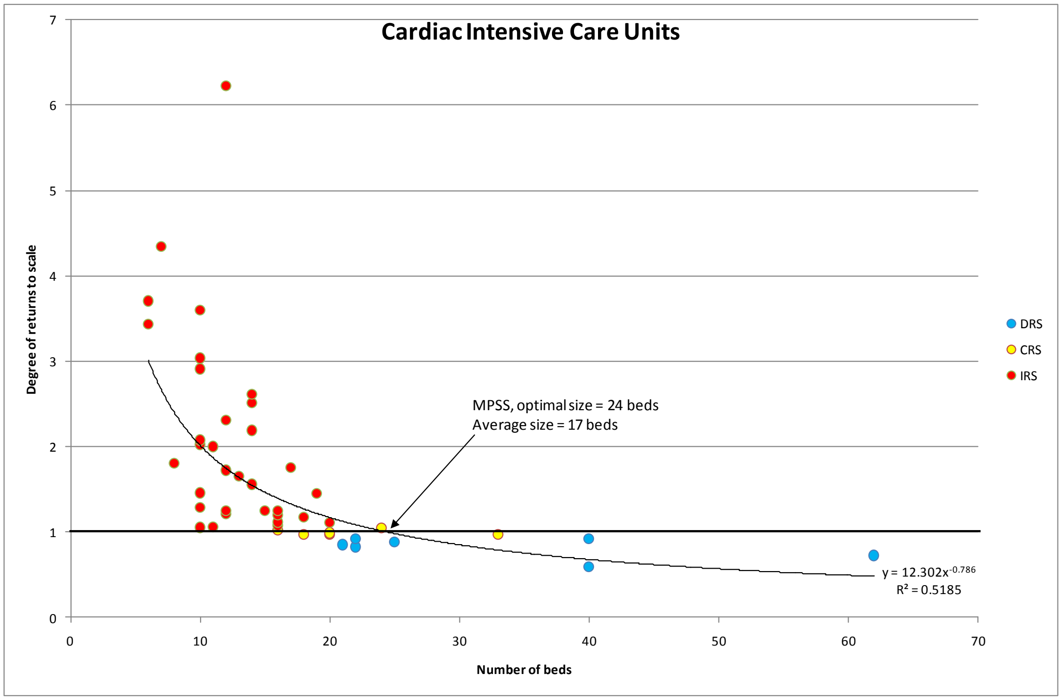
Table 2. Estimation of returns to scale for intensive care units

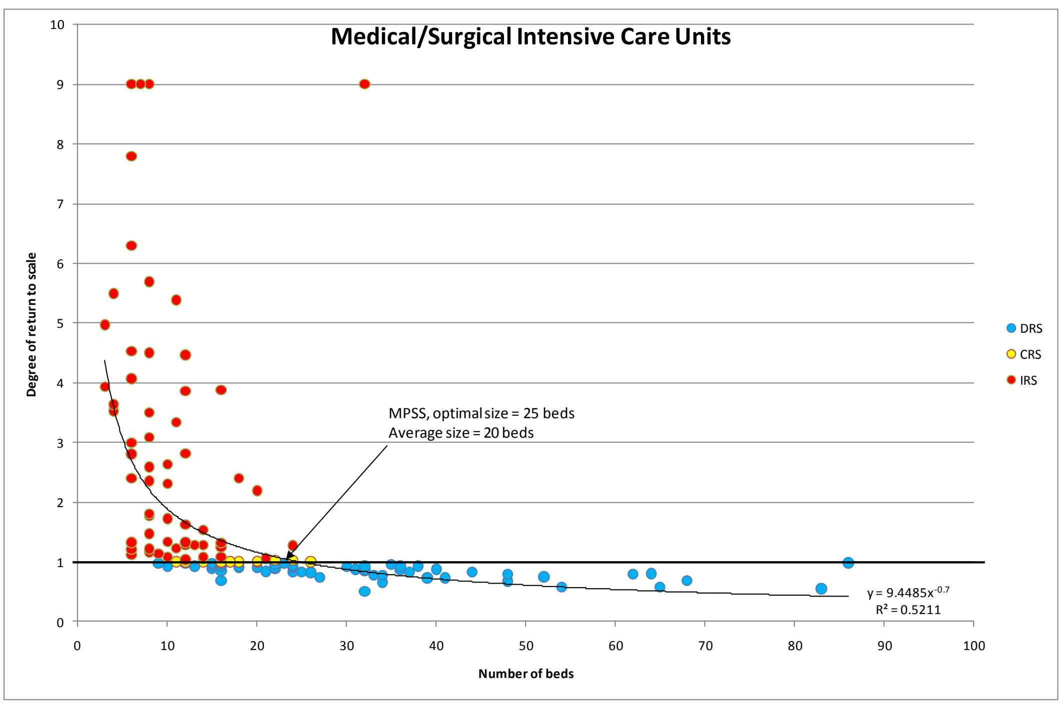
	DRS	*CRS*	*IRS*	*Total*
# of CICUs	7	6	35	48
% of total	15%	13%	73%	100%
Average alpha	0,82	1,00	2,02	1,72
Average # beds	33	22	12	17
Average nurses hours	529	383	240	297
Nurses hours / bed	16.0	17.5	19.3	17.9
# of MSICUs	65	14	62	141
% of total	46%	10%	44%	100%
Average alpha	0.85	1.00	3.12	0.81
Average # beds	30	18	10	20
Average nurses hours	535	315	151	344
Nurses hours / bed	17.8	17.5	14.7	17.0
# of PICUs	4	3	18	25
% of total	16%	12%	72%	100%
Average alpha	0.68	1.0	2.32	2.00
Average # beds	21	18	9	12
Average nurses hours	435	345	141	212
Nurses hours / bed	20.7	18.8	16.6	18.2
# of NICUs	6	3	42	51
% of total	12%	6%	82%	100%
Average alpha	0.77	1.01	1.75	1.59
Average # beds	44	38	18	22
Average nurses hours	878	394	276	353
Nurses hours / bed	19.9	10.4	15.6	16.1

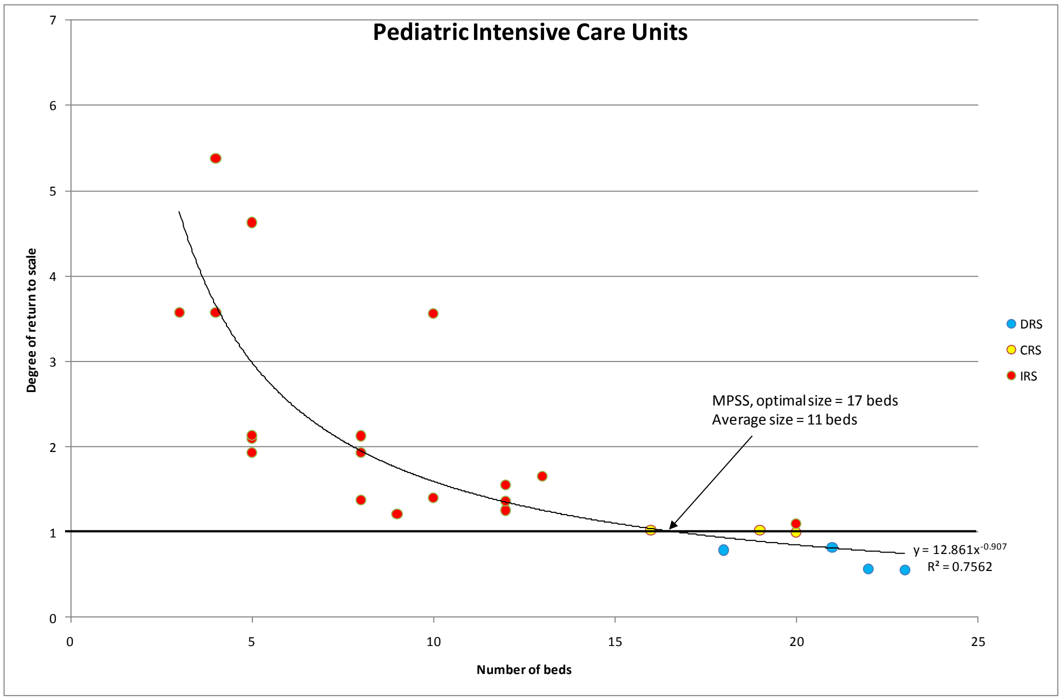
The primary finding is that intensive care units are mostly operating under increasing returns to scale: 73% of CICUs, 44% of MSICUs, 72% of PICUs and 82% of NICUs. For the α's interpretation, recall in the method section, the production technology T is said to be *homogenous of degree* α if for all $\beta > 0$: $(x, y) \in T \Rightarrow (\beta x, \beta^{\alpha} y) \in T$. Therefore we can interpret the α's as the elasticity of the output regarding increase in the intensive care units' inputs. While most of intensive care units in IRS show a degree of homogeneity near 2, we can conclude that on average an increase in 1% in units inputs will lead to an increase in 2% in the output, demonstrating a potential substantial gain in productivity.

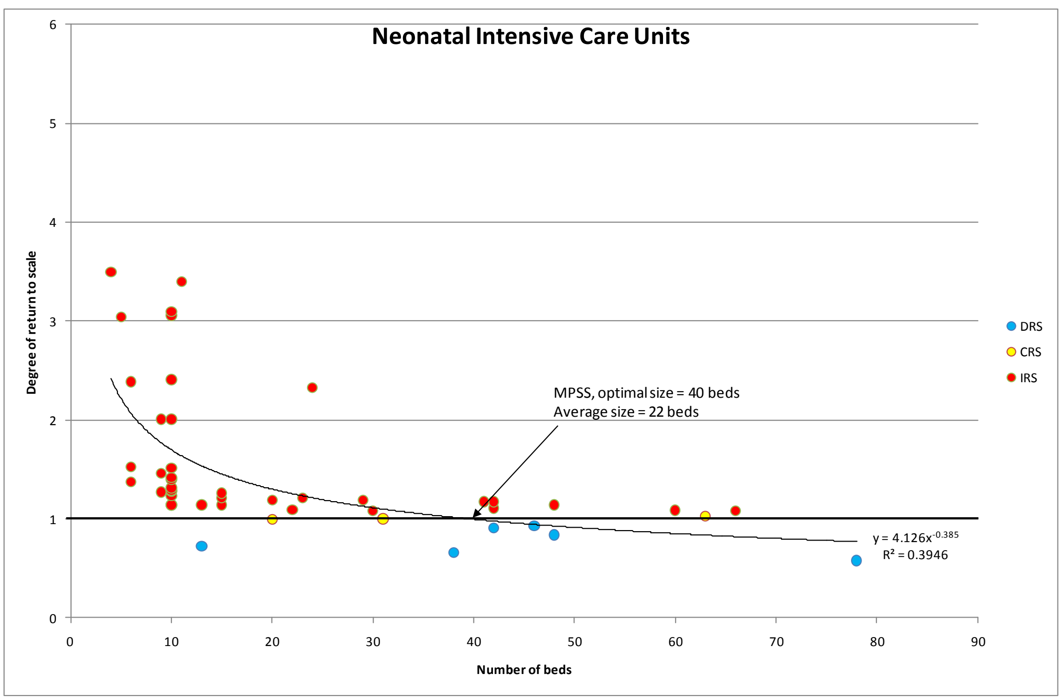
The relationship between the degree of returns to scale and the bed size of each intensive care units are presented in Figure 1.

Figure 1. Returns to scale and bed size of intensive care units









For all types of ICUs we find a decreasing and concave relationship between the degree of returns to scale and the bed size. Units with an estimated degree of homogeneity located above one are operating under increasing returns to scale (red points in the figure), while decreasing returns to scale are characterized by blue points and CRS by yellow points. We clearly see the predominance of increasing returns in all four intensive care units. The optimal productivity size of the units is given by the number of beds where the returns to scale curve crosses the CRS line at α=1. For CICUs the optimal productivity size is 24 beds while the observed average size is 17 beds. Therefore, units in cardiac intensive care should increase their size by an average of 7 beds. We have the same diagnosis for medical/surgical and pediatric ICUs. For neonatal care the increase in size is more pronounced (40 beds at optimal versus 22 beds for observed average). We reiterate here that since we find concavity in the relationship between number of beds and scale, these findings would have been obscured if determining returns to scale using a strictly convex production frontier.

We also ran an econometric analysis demonstrating the elasticity of the degree of returns to scale to the number of beds. These findings are summarized in Table 3.

Table 3. Elasticity of the degree of returns to scale to the number of beds
– Model: RTS = C * Beds$^{\beta}$ –

	Elasticity β	*Student t*	*R²*	*Average beds*	*Optimal beds*
CICUs	-0.79	7.04	0.52	17	24
MSICUs	-0.70	12.30	0.52	20	25
PICUs	-0.91	8.45	0.76	11	17
NICUs	-0.39	5.65	0.39	22	40

We can interpret these elasticities as the “natural speed of convergence or divergence” towards or from the optimal scale depending on the starting position (IRS or DRS). Since most of intensive care units are under increasing returns to scale, they have to increase their size in average to reach constant returns to scale and the optimal productivity size. The elasticity results indicate that the “return” of each percent increase in the number of beds is better in pediatric, cardiac and medical/surgical intensive care units rather than in neonatal intensive care units: the convergence toward the optimal scale is faster for the former. Therefore, starting from increasing returns to scale, intensive care units converge quickly towards the CRS scale by increasing the number of beds and they depart more slowly from optimal scale when decreasing the number of beds.

In order to maintain a level of appropriate treatment, other inputs, particularly nurses must also be included in any study of ICU scale efficiency. Including FTE nurses hours, we found that nursing hours need to increase by 43% in NICUs, 60% in CICUs, double the number in MSICUs and more than double for PICUs. Because the nursing hours are more variable vis-à-vis bed numbers, production technologies also vary. The finding in the PICUs which demonstrate the greatest increase in nurses, may be due to the higher level of teaching activities (number of medical residents) in this unit as compared to the others because of the explicit accounting for medical education, another output.

Interestingly, when assessing inputs separately we arrived at varying results. Looking at the average number of nursing hours/beds ratios, we see that they are all close to the optimal ratio except for neonatal units. In this case, the number of beds should be halved (moving from DRS to CRS) resulting in a decrease in the ratio of 19.9 at DRS nurses hours/beds compared to the optimal value of 10.4 at CRS scale.

Results for hospitals

We now present our results for hospitals as a whole. In addition to the inputs and output specified by unit, we add a case mix indicator in order to control for patient severity since unlike the specific units, hospitals are heterogeneous providing multiple outputs with multiple inputs. Results concerning returns to scale are provided in table 4.

Table 4. Estimation of returns to scale for hospitals

	DRS	*CRS*	*IRS*	*Total*
# of Hospitals	112	24	24	160
% of total	70%	15%	15%	100%
Average alpha	0.71	1.0	2.02	0.98
Average # beds	418	140	68	324
Average nurses hours	5000	2016	839	3928
Nurses hours / bed	12.0	14.4	12.4	12.1

Recall that we found overwhelming support that most of the individual intensive care unit exhibited IRS (60%), however, the majority of hospitals operate under DRS (70%). For hospitals, the average number of beds at CRS is 140 while observed size is 324 for hospitals on average. The optimal size should therefore be about half of the observed size. Similarly, we find in order to move from DRS to CRS, the number of nursing hours needs to be more than halved. Interestingly since the decrease in beds and number of nursing hours is of the same order, the input mix ostensibly remains the same.

Given that arguments in the past have included ownership as a determinant of hospital economic behavior, we likewise test for the relationship between ownership and returns to scale at the hospital level (Table 5).

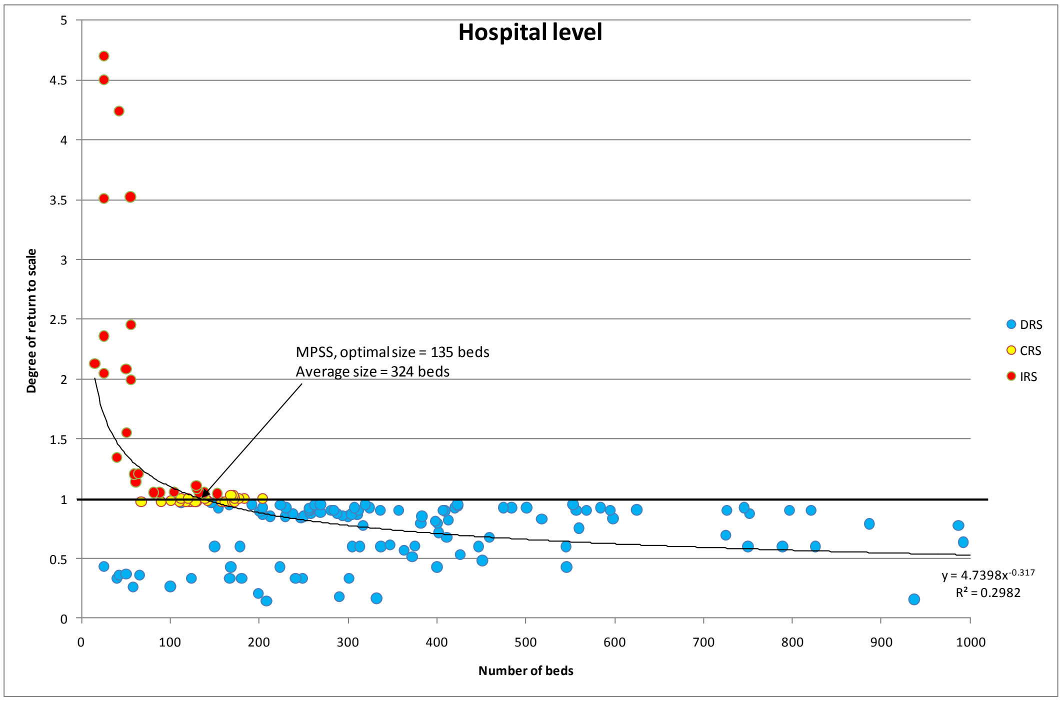
Table 5. Returns to scale and ownership for hospitals

Returns to scale / *Ownership*	*DRS*	*CRS*	*IRS*	***Total***
Public (Government owned)	10	4	4	**18**
For-Profit	42	9	8	**59**
Not-for-Profit	60	11	12	**83**
Total	**112**	**24**	**24**	**160**

Chi² test of independence: $p = 0.71$. Accept H0: independence.

We find no statistically significant relationship between returns to scale and the ownership form, indicating that returns to scale and number of beds does not reflect any type of organization mission. This 'none' finding is interesting since public hospitals typically need reservation quality (extra beds) for emergency purposes as the safety net in the community; and Not-for-profit hospitals may have extra beds so that a paying patient is not turned away. It is also interesting, that For-Profit hospitals do not appear to be minimizing costs by operating at CRS, especially because part of their mission is to earn dividends for the shareholders.

Figure 2. Returns to scale and bed size of hospitals



Given the results in Figure 2, we find that the estimated optimal productivity size is about half of the average number of beds for all hospitals.

Discussion and conclusion

In this paper we empirically apply the Boussemart et al.[7,22] approach of measuring returns to scale for multi-output homogenous technology to a population of intensive care units and hospitals operating in Florida during 2005. We pursued this course of action on the suggestions by Wilson and Carey [10] and Burgess [18] who suggest that non-parametric methods be used for ascertaining hospital returns of scale but extend beyond the typical DEA approaches that require the convexity assumption. Applying the α–returns to scale approach, we found that for all four types of intensive care units: CICUs, MSICUs, PICUs, and NICUs, there exists overwhelming evidence of increasing returns to scale. We find contrasting results at the hospital level where most of hospitals appear to operate under DRS indicating that they are too large in terms of bed size. The optimal productive size was determined to exist at 140 beds whereas the average size of hospitals was 324. We did not find any evidence that property rights theory applied at the hospital level. Therefore, we cannot make any assertions that deviating from CRS has to do with any organizational mission – such as reservation quality (slack), providing services for all in need, or profit maximization.

There exist several reasons as to why the CICUs, PICUs, and NICUs, in our sample, operate at increasing returns to scale. First, they are more specialized requiring highly specialized inputs including specially trained nurses and beds which are very expensive. Second, PICUs are usually smaller since those patients aged younger than 18 are not typically severely ill and therefore the demand for this type of intensive care may not be as great for hospitals in general to increase its size. Third, NICUs are also highly specialized but do not need the capital intensity that other ICUs require since these units are used for premature babies requiring support. Conversely, the MSICUs are more generalized and beds maybe substituted between the intensive care unit and the general medical/surgical wards. Hospital managers may also opt to keep intensive care units small since they are very expensive, in an accounting sense.

However, we are assessing the economics of these services, which given our findings suggest that the units, especially the CICUs, PICUs, and NICUs are too small. Therefore, one hospital policy/managerial decision may include reallocating resources from hospital general beds to the more specialized ICUs' beds, which could be one response to changes hoped for in the new health care reforms enacted by the US Congress and the Obama Administration. In the best laid plans put forth in the new health care reform legislation, the medical focus would shift towards more preventive care which should be complemented by a decrease in general inpatient hospital services (especially chronic diseases such as diabetes, asthma, complications from obesity, for example). However, despite any attempt towards preventive/primary care may not have the same result particularly demand/need for more intensive care beds since these services would be needed irrespective of health care reform provisions, at least in the shorter run. Another consequence of a reallocation of resources from hospital level to ICUs is on quality of care. If ICUs were able to increase volume of patients along with increased bed sizes and the requisite number of nurses, then an argument can be made that quality would likewise increase in these larger more practiced environments.

References

1. Edbrooke, D., Hibbert, C., Ridley, S. Long, T., Dickie, H., and the Intensive Care Working Group on Costing (1999).” The development of a method for comparative costing of individual intensive care units” *Anaesthiesia* 54:110-120.

2. Jacobs, P., Rapoport, J., and Edbrooke, D.,(2004) “Economies of scale in British intensive care units and combined intensive care/high dependency units” *Intensive Care Medicine*. 30:660-664.

3. Jones, J., Rowan, K. (1995) “Is there a relationship between the volume of work carried out in intensive care units and its outcomes?” *International Journal of Technical Assessment in Health Care*. 11(4):762-769.

4. Vita, M. (1990). “Exploring hospital production relationships with flexible functional forms” *Journal of Health Economics*. 9:1-21.

5. Vitalino, D. (1987). “On the estimation of hospital cost functions” *Journal of Health Economics.* 16:305-318.

6. Dranove, D. (1998).”Economies of scale in non-revenue producing cost centers: Implications for hospital mergers” *Journal of Health Economics*. 17:69-83.

7. Boussemart, J-P., Briec, W., Peypoch, N., Tavéra, C. (2009) “α-Returns to scale and multi-output production technologies” *European Journal of Operations Research,* 197(1): 332-339.

8. Yafchak, R.(2000) “A longitudinal study of economies of scale in the hospital industry” *Journal of Health Care Finance* 27(1): 67-89.

9. Magnussen, J. (1996) “Efficiency measurement and the operationalization of hospital production” *Health Services Research*. 31(1):21-37.

10. Wilson, P. and Carey, K. (2004) “Nonparametric analysis of returns to scale in the US hospital industry” *Journal of Applied Econometrics*. 19:505-524.

11. Preyra, C. and Pink, G. (2006). “Scale and scope efficiencies through hospital consolidations” *Journal of Health Economics* 25:1049-1068.

12. Carey, K. (1997) “A panel data design for estimation of hospital cost functions” *Review of Economics and Statistics,* 79(3): 443-453.

13. Lindrooth, R., LoSasso, A., and Bazzoli, G. (2003) “The effect of urban hospital closure on markets” *Journal of Health Economics*. 22(5): 691-712.

14. Joskow, P. (1980). “The effects of competition and regulation on hospital bed supply and the reservation quality of hospital” *Bell Journal of Economics The Rand Corporation.*11 (1):377-398.

15. Connor, R., Feldman, R., Dowd, B., Radcliffe, T. (1997). “Which type of hospital mergers save consumers money?” *Health Affairs,* 16(6): 52-74.

16. Hollingsworth, B. (2008) “The measurement of efficiency and productivity in health care delivery” *Health Economics*. 17:1107-1128.

17. Valdmanis, V. (2010) “Measuring economies of scale at the individual hospital and the city-market level” *Journal of Health Care Finance.* Forthcoming.

18. Burgess, J. (2006).”Productivity analysis in health care” *The Elgar Companion to Health Economics* ed. By Andrew Jones, Edward Elgar, Chetenham.

19. Gyldmark, M. (1995) “A review of cost studies of intensive care units: Problems with the cost concept” *Critical Care Medicine.* 23: 964-972.

20. Tsekouras, K., Papathanassopoulos, F., Kounetas, K., Pappous, G. (2010). “Does the adoption of new technology boost productive efficiency in the public sector? The case of ICUs system” *International Journal of Production Economics* 128:427-433

21. Lau, L.J. (1978). “Application of profit functions”. In: Fus, McFadden (Eds.), Production Economics: A Dual Approach to Theory and Applications. North-Holland, Amsterdam.

22. Färe, R., Mitchell, T. (1993) “Multiple outputs and homotheticity” *Southern Economic Journal*. 60: 287–296.

23. Boussemart, J-P, Briec, W. Leleu H. (2010). “Linear programming solutions and distance functions under α-returns to scale” *Journal of the Operations Research Society*. doi:10.1057/jors.2009.97. (Forthcoming).

24. Charnes, A., Cooper, W.W., Rhodes, E. (1978) "Measuring the efficiency of decision making units" *European Journal of Operational Research*. 2: 429-444.

25. Banker, R., Charnes, A., Cooper, W. (1984) "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis" *Management Science*. 30(9): 1078 1092.

26. Färe, R., Grosskopf, S., Njinkeu, D. (1988) "On piecewise reference technologies" *Management Science.* 34: 1507-1511.